DOCUMENT RESUME

ED 166 134 95 SP 013 572

AUTHOR Clark, Christopher M.; Yinger, Robert J.
TITLE Methods for Discovering Cues Used by Judges: Two Working Papers. Research Series No. 11.
INSTITUTION Michigan State Univ., East Lansing. Inst. for Research on Teaching.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE Apr 78
CONTRACT 400-76-0073
NOTE 14p.
AVAILABLE FROM Institute for Research on Teaching, College of Education, Michigan State University, 252 Erickson Hall, East Lansing, Michigan 48824 ($1.00)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Cues; Educational Research; Literature Reviews; Policy; Research Criteria; *Research Methodology; Research Tools; *Statistical Analysis; *Teacher Behavior; *Teacher Characteristics; Teaching Methods
IDENTIFIERS *Teacher Judgment

ABSTRACT

Two research papers focusing on approaches and methods for discovering cues used by teachers in making judgments in the classroom are presented here. The approaches described in the first paper are: review of empirical and theoretical literature concerning the objects to be judged and the behavior of judges; interview of judges to determine what they believe the salient cues are; participant observation of situations in which the judgments of interest are taking place; and choice of very simple objects to be judged. The quantitative methods for discovering cues are listed in the second paper as follows: prespecify a large list of potentially relevant cues and then use observation and regression analysis to narrow the list; use the semantic differential; use multidimensional scaling; and use Kelly's Role Repertory Test. References are included. (DS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED166134

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Robert J. Yinger

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Research Series No. 11

METHODS FOR DISCOVERING CUES USED BY JUDGES:

TWO WORKING PAPERS

Christopher M. Clark and Robert J. Yinger

Published By

The Institute for Research on Teaching
252 Erickson Hall
Michigan State University
East Lansing, Michigan 48824

April 1978

The work reported herein is sponsored by the Institute for Research on Teaching, College of Education, Michigan State University. The Institute for Research on Teaching is funded primarily by the Teaching Division of the National Institute of Education, United States Department of Health, Education, and Welfare. The opinions expressed in this publication do not necessarily reflect the position, policy, or endoresement of the National Institute of Education. (Contract No. 400-76-0073)

SP013572

Contents

Approaches to Identifying Cues for Policy-Capturing Research 1

by Christopher M. Clark

Review of Theoretical and Empirical Literature 1

Self-Report of Judges. 2

Participant-Observation of the Judgment Situation. 3

Use of Very Simple Objects to be Judged 3

Four Quantitative Methods for Discovering Cues Used by Judges. 5

by Robert J. Yinger

Pre-specifying Cues. 5

The Semantic Differential. 7

Multidimensional Scaling 7

Kelly's Role Repertory Test. 8

References 11

METHODS FOR DISCOVERING CUES USED BY JUDGES:
TWO WORKING PAPERS

Approaches to Identifying Cues for Policy-Capturing Research

by Christopher M. Clark*

The study of teacher judgment is one of the major concerns of the Institute for Research on Teaching. One powerful method of studying and representing human judgment is policy capturing (see for example, Rappoport & Summers, 1973). Unfortunately, the literature describing the methodology of policy capturing does not provide much guidance on ways to identify and select cues (or features) of the objects to be judged.

Reflection on this problem has led to consideration of four alternative ways of generating a cue list for policy-capturing studies. The four approaches are: (1) review of empirical and theoretical literature concerning the objects to be judged and the behavior of judges, (2) interview of judges to determine what they believe the salient cues are, (3) participant observation of situations in which the judgments of interest are taking place, and (4) choice of very simple objects to be judged. Each of these approaches will be discussed below.

Review of Theoretical and Empirical Literature

This approach involves two possible foci. If the literature identifies those cues or features of objects to be judged that are usually used by judges, then such literature constitutes a set of nominations for cues to be employed in future studies. If hardly any such literature exists, as is the

---

*Christopher M. Clark, senior researcher at the Institute for Research on Teaching and assistant professor of educational psychology, is coordinating a study of teacher planning.

case in research on teacher judgment, then it would be most profitable to examine literature that deals with the objects to be judged and their functional relationships to desirable outcomes. An example of the latter approach is Anderson's study (1977), where the literature on teacher effects was reviewed to generate a list of cues, characteristics of teachers that correlate with student achievement. Anderson used these cues to systematically vary descriptions of effective and ineffective teachers; these descriptions were then judged as effective or ineffective by experienced high school teachers.

Self-Report of Judges

The second approach to generating a cue list involves asking judges (in our case, teachers) to identify the important features or cues they think influence their judgments about the objects to me judged. This information can then be summarized, abstracted, and used to create sets of objects to be judged that vary systematically on the reportedly significant cues. A recent IRT study by Clark, Wildfong, and Yinger (report in preparation) exemplifies this approach. Thirteen experienced teachers were asked to rate the attractiveness of language arts activities. After rating the activities, the teachers were asked to reexamine each activity description they rated high and list the features of that activity that contributed to their judgment. This process was repeated for activity descriptions rated low. All of the features identified by the teachers were sorted by the experimenters into categories which constituted a pool of potential cues, features, or dimensions on which objects to be judged (in this case, descriptions of language arts teaching activities) might be varied in future policy-capturing studies.

Participant Observation of the Judgment Situation

The third approach involves careful observation of naturally-occurring instances of the judgments of interest. Here, the observer becomes immersed in the judgment situation, attempting to understand the judge's frame of reference and, indeed, becomes a judge. Identifying important cues or features of the objects to be judged then becomes a matter of introspection. The participant-observation approach can, of course, be combined with the second approach described above; the important distinction is that the interrogator in this case is much more intimately familiar with the judgment situation than is the interrogator in the "self-report of judges" approach.

Use of Very Simple Objects to be Judged

The final approach to generating a cue list involves choosing objects to be judged in such a way that an exhaustive list of their features will be sufficienctly short to permit full factorial experimental manipulation. An example of this approach can be seen in a study by Hammond and Adelman (1976) in which the objects to be judged were types of bullets being considered for use by the Denver Police Department. The bullets varied on only three features. It may be that policy-capturing approaches are most useful when the judgments in question are concerned with relatively simple objects.

Using this approach for our purposes, however, generates the question, "Under what circumstances do teachers exercise judgment of objects that involve only four or five different features?" It may be that our use of policy-capturing methodology should be limited to such situations; we may

be able to learn much about teacher judgment in general by examining a few simple examples. Just as experimental psychologists have learned much about human memory and information processing by studying human performance in remembering nonsense syllables, so we too might be able to generalize beyond the simple judgment situations observed to basic processes in the mental life of the teacher.

Four Quantitative Methods For Discovering Cues Used by Judges

by Robert J. Yinger*

Wilcox (1972) discusses four quantitative methods for discovering cues used by judges in a policy-capturing study. The first is to pre-specify a large list of potentially relevant cues and then use observation and regression analysis to narrow the list. The second is to use the semantic differential. The third involves the use of multidimensional scaling, and the fourth is the use of Kelly's Role Repertory Test. Each of these methods will be discussed in turn.

Pre-specifying Cues

The first method, that of pre-specifying a large list of potentially relevant cues and then narrowing them down using regression analysis is exemplified by Slovic's (1969) study of stockbrokers' decision making. Through discussions with two stockbrokers, Slovic identified 11 cues he felt were potentially important to them in their decision making. After varying these cues factorially, he had the stockbrokers make decisions on the resulting stock profiles. The two were then each given resulting 128 standardized descriptions of the stocks and asked to rate them on a 9-point preference scale.

An interesting finding of the study was that neither decision maker appeared to be using more than half of the available attributes. This finding raises a question about whether or not Slovic initially listed all the

---

*Robert J. Yinger, IRT researcher, is collaborating with Christopher M. Clark in a study of teacher planning.

attributes that were relevant to the decisions. Wilcox, commenting on the study, mentions that there is no way to judge from Slovic's results whether or not he included too few relevant cues; the method used provides no corrective signals when not enough relevant attributes have been included. According to Wilcox, Slovic utilized artificial alternatives that did not correspond to known real alternatives. The only information about the hypothetical stocks available to the decision makers was in terms of the 11 attributes selected by Slovic. Therefore, it is not surprising that the decision makers' preferences were highly correlated with the given attributes. One way to test whether or not the cues were really being used would be to have the stockbrokers rate stocks with which they were already familiar. In this way, the cues would emerge from a real life judgment task rather than an artificial one. Any sharp reduction in the explanatory power of the model would provide a signal of possible mis-specification of relevant attributes.

Two disadvantages are evident in Slovic's method. First, the method is relatively inefficient. Since Slovic had to cross factorially all of the relevant cues, the tasks became so difficult as to require the judges to spend about 10 hours on each task. Second, the tendency for the observer to project his own perceptions of the cues onto the task is inherent in the methodology of the first step, where the observer, via observation and discussion of cues with the judges, decides which cues appear to be relevant to the situation.

The Semantic Differential

The second quantitative method that has been widely discussed for discovering cues is the semantic differential. This method is based on factor analysis of multiple judges' ratings of objects on a large number of pre-specified adjective scales. The objects are then rated on the summarizing attribute factors revealed by factor analysis, which in turn is based on ratings by all judges on all scales. By comparing the factor scores of the objects to an individual judge's rating, a measure is obtained of each individual's assumptions (cues they are using).

This method is more efficient than using regression analysis alone, mainly due to the reduction of the number of factors that a judge has to deal with. The major disadvantage, however, is that individual differences are obscured by the combination of many judges' ratings prior to the factor analysis. It is also very difficult to make pre-specified adjective scales that are relevant to the problem at hand. A third disadvantage of such an approach is that strong assumptions are made about the metric nature of the scales by having the judges rate the objects on a standard 7-or 9-point scale.

Multidimensional Scaling

The third quantitative method used for discovering what cues a judge is using, multidimensional scaling, uses estimates or comparisons of inter-object similarities to build up a spatial configuration of objects in which similarities correspond to inter-object distances. This configuration is then analyzed, and the minimal number of dimensions by which the configuration may be embedded is determined. (See Wiggins, 1973, for more discussion

of this method and several examples of its use.)

There are two primary advantages of the multidimensional scaling method. First, it is not necessary to pre-specify the cues being used by the judge since cues are elicited during the similarity-comparison task. Second, only weak ordinal assumptions regarding the types of comparisons of similarity used by the judges are required.

Three major disadvantages of this method are discussed by Wilcox. First, it requires large numbers of similarity comparisons in order to construct the stable metric of the cue dimensions. This is due to the necessity of making all possible triad comparisons using the objects selected for the task. Second, comparisons of triads often requires extra work for the judge in terms of calculating combinatorial weights. Since a triad comparison is comparing two objects against a third, the two objects must somehow be weighted to determine their importance. Finally, it is difficult to interpret the dimensions arrived at after the task since introspective information is not elicited, i.e., cue labels are not asked for.

Kelly's Role Repertory Test

Kelly's Role Repertory Test (1955), the fourth quantitative method for arriving at cue dimensions, generally involves four steps. First, the judge is asked to match a given list of appropriate "role" descriptions with appropriate objects from his or her own experience. Next, a limited number of triads of these objects are selected, and the judge is asked which pair of the triad is most similar, in what ways they are so, and in what important ways the third member of the triad differs. Then, the judge positions each object on each relevant attribute scale. Objects are scored as either

+1 (similar) or -1 (different) on each of the raw cue dimensions implicitly defined in the first step. Finally, these attribute data are factor analyzed to eliminate redundancies.

An advantage of using this method is that the comparisons task is simpler for the judge than that of multidimensional scaling because individualized, self-selected familiar objects are used. An additional advantage of this method is that attribute lables used by the judge are elicited during the task. After these labels are elicited, the similarity comparison task is cut short, further reducing task difficulty for the judge.

The method used by Wilcox (1972) in his stock market participant study is basically a variation of Kelly's Role Repertory Test. Like Kelly's method, a two-stage data collection procedure was used. In the first stage, called the Stock Role Repertory Exercise, a list of 20 roles that various stocks play in the subject's experience and conceptual structure was prepared (e.g., "a very popular stock," "the stock in which he first made a considerable gain," "the stock sold too soon". Next each judge was asked to designate a particular stock for each of these "roles." Twenty triads of these stocks were then selected and presented to the judges, who were asked in what important way two members of the triad differed from the third member (as in Kelly's triad comparison task). This step was used to elicit important conceptual dimensions used by the judge. (It is at this point in the method that considerable experimenter judgment is called for. Wilcox reported a 30% reduction of labels made at this point as he tried to eliminate redundancies.) After the relevant cues had been determined, a questionnaire was constructed which asked the judge to do the following for each attribute elicited in the previous step:

1. Divide the scale into 2 to 9 equivalent intervals.
2. Place any appropriate stocks into two separately provided categories - "scale not relevant" or "not enough information."
3. Place the remaining stocks on the attribute scale at their appropriate intervals.

The second step of Wilcox's method involved a factor analysis (using principal component anaysis) of the questionnaire data to condense the "raw attributes" into "attribute factor structures." The factor analysis, using data for a single decision maker, eliminates most of the previously-mentioned difficulties of the semantic differential. Also, by allowing the judge to divide the scale into between 2 and 9 equivalent intervals, much more ordinal or metric information is provided on each attribute scale. This is in contrast to Kelly's method and to the multidimensional scaling method which both make the assumption of equal scale intervals for all attributes.

Two disadvantages of this method are most apparent. First, once the factor analysis has been completed, the factors themselves are left unnamed since the data are only used for predictive purposes. Thus, the simplified factor structures do not really offer summary value for descriptive purposes. Furthermore, it is relatively easy to reliably measure the relevant assumptions of a judge who is dealing with a large number of fairly simple-consequence judgments made in the same domain, but it is much harder when a judge is dealing with only a very few large decisions. It is harder still when these decisions have very complicated action spaces. This drawback mat be inherent to all the methods described; they can be efficiently and easily used only on simple judgment tasks. It may be that the extrapolation of these methods to much more complicated task environments such as teacher decision making may be very difficult and, in some cases, impossible.

References

Anderson, B. Differences in teachers' judgment policies for varying numbers of verbal and numerical cues. Organizational Behavior and Human Performance, 1977, 19, 68-88.

Hammond, K.R., & Adelman, L. Science, values, and human judgment. Science, October 1976, 194, 389-396.

Kelly, G.A. The psychology of personal constructs: A theory of personality (Vol. 1). New York, W.W. Norton & Company, Inc., 1955.

Rappoport, L., & Summers, D.A. Human judgment and social interaction. New York: Holt, Rinehart and Winston, 1973.

Slovic, P. Analyzing the expert judge. Journal of Applied Psychology, 1969, 53, 255-263.

Wiggins, N. Individual differences in human judgments: A multivariate approach. In L. Rappoport & D. Summers (Eds.), Human judgment and social interaction. New York: Holt, Rinehart and Winston, 1973.

Wilcox, J. A method for measuring decision assumptions. Cambridge: MIT Press, 1972.